https://farid.ps/articles/industrialization_of_profiling/en.html

From Persons of Interest to Populations: The Industrialization of Profiling

The emergence of profiling—from targeted, manual scrutiny of specific "persons of interest" to the automated, continuous monitoring of entire populations—represents one of the most profound transformations in the exercise of power, the role of technology, and the boundaries of individual autonomy. What once required significant human effort, institutional prioritization, and deliberate selection has evolved into a seamless infrastructure that generates, aggregates, and analyzes behavioral data on billions of people in real time, often as an incidental byproduct of everyday life.

This transformation was not predetermined by technology alone. It emerged from the interaction of bureaucratic expansion, repeated security crises, economic incentives tied to data monetization, and the relentless reduction in the marginal cost of data collection, storage, and inference. The result is not simply "more surveillance," but a qualitatively different regime: one that replaces natural friction with frictionless scale, human discretion with algorithmic automation, and exceptional suspicion of the few with baseline observation of the many.

At its core lies a fundamental metamorphosis: profiling has shifted from an **artisanal craft**—selective, labor-intensive, and explanatory—to an **industrial process**—universal, automated, and predictive. What follows traces that transformation, identifying the moments where constraints eroded and new capabilities crystallized into a system of continuous, population-wide inference.

I. Foundations: Profiling as Selective, Manual Practice

Profiling, in its most basic form, involves the systematic collection and interpretation of information to infer characteristics, predict behavior, or assign categories of risk. Its origins extend deep into antiquity.

Ancient empires conducted censuses not merely for taxation or conscription, but for classification. Roman authorities and Chinese imperial administrators sorted populations by occupation, loyalty, and status, producing early relational maps that could identify potential threats. Religious institutions maintained records of births, marriages, confessions, and moral conduct, constructing proto-social graphs that revealed networks of influence and deviation.

Yet these systems shared a defining constraint: **information was expensive**. Gathering, verifying, storing, and interpreting data required significant human labor. As a result, profiling remained **selective, episodic, and bounded**. It focused on elites, dissidents, or strategically relevant groups—not entire populations.

In early modern Europe, this selectivity persisted even as states expanded their surveillance apparatus. Intelligence efforts targeted heretics, political rivals, smugglers, and foreign agents through informants, intercepted correspondence, and physical surveillance. The *cabinets noirs*—or Black Chambers—of France and other states epitomized this approach: teams of clerks manually opened letters, copied them, and resealed them for delivery. These operations were inherently constrained. They focused on high-value targets because anything broader was logistically impossible.

Even at this stage, however, the power of **metadata** was clearly understood. Information about communication—sender, recipient, timing, and route—could expose networks and intentions without requiring access to content. The 1844 British Post Office espionage scandal brought this into sharp public focus. Italian revolutionary Giuseppe Mazzini, an exile in London, suspected his letters were being opened by authorities at the request of foreign powers. He and his supporters placed poppy seeds and grains of sand inside envelopes as markers; when the letters arrived disturbed, Mazzini prompted radical MP Thomas Duncombe to raise the issue in Parliament. The ensuing scandal revealed systematic letter-opening under warrants issued by Home Secretary Sir James Graham, sparking outrage, parliamentary inquiries, and the eventual abolition of the Post Office's secret department. It marked one of the first modern privacy panics and underscored how relational data alone could dismantle networks of association.

In response, legal norms such as the "secrecy of correspondence" (*Briefgeheimnis*, *secret de la correspondance*) emerged. These principles restricted the use of communication data strictly to operational purposes like delivery, prohibiting secondary exploitation for surveillance or profiling. The underlying idea was simple but profound:

> Data generated for a specific function should not be repurposed to construct broader profiles of individuals or networks.

This principle would echo across centuries—yet ultimately erode under technological and institutional pressure.

II. The Bureaucratic Century: Scaling Without Automation

The 20th century expanded profiling dramatically while preserving many of its earlier constraints. The demands of total war required unprecedented information gathering. Mail censorship, signals intelligence, and codebreaking extended surveillance beyond elites to wider populations. Institutions such as the National Security Agency institutionalized large-scale interception, while domestic agencies compiled extensive files on political groups, suspected radicals, and criminal networks.

Yet profiling remained **fundamentally targeted**. Wiretaps were tied to specific individuals or lines. Intelligence files were curated by human analysts. Even as volume increased, **human attention remained the bottleneck**.

Early computing systems (1950s–1970s) began to change the scale of record-keeping. Governments and corporations digitized welfare rolls, credit histories, and criminal databases, enabling faster retrieval and cross-referencing. But these systems still operated on **discrete records**, not continuous streams of behavior.

By the 1970s, concerns about centralized "data banks" prompted legal responses. The U.S. Privacy Act of 1974 and early European data protection laws introduced principles of purpose limitation, data minimization, and transparency. These frameworks extended the logic of correspondence secrecy into the digital age.

However, they were built on a crucial assumption: that data collection was **bounded and episodic**. They regulated records—not flows. This assumption would soon collapse.

III. The Inflection Point: From Records to Data Exhaust

The decisive break occurs in the late 1990s and early 2000s with the rise of the internet—not merely as a communication medium, but as an infrastructure that continuously produces data.

Digital systems generate **data exhaust**: metadata created automatically as a byproduct of ordinary activity. Every connection, query, click, and movement produces traces that can be logged, stored, and analyzed at negligible cost.

This marks the decisive shift:

> Profiling ceases to be an activity performed on data and becomes an infrastructure that continuously produces it.

Internet Service Providers capture connection logs, DNS queries, and routing information, revealing patterns of behavior even without content access. Unlike postal metadata—ephemeral and decentralized—digital metadata is persistent, centralized, and trivially searchable.

On top of this infrastructure, platforms such as Google and Meta transformed profiling into a core economic model. Search engines capture intent; social networks map relationships; mobile ecosystems track movement. Embedded trackers extend visibility across vast portions of the web. Meta's tracking pixels, present on roughly one-third of the world's popular websites, monitor activity far beyond its own platforms, often capturing sensitive signals from health, finance, or political contexts.

A critical realization emerges in this environment:

> Content becomes largely redundant. In many cases, relational patterns are not merely proxies for meaning—they are more analytically useful than content itself.

Metadata does not simply indicate that communication occurred; it enables **probabilistic reconstruction of content**. Who communicates with whom, when, how often, and within

what broader context can strongly constrain what is being communicated. Publicly available information—shared affiliations, professional roles, political positions, social ties—further narrows the space of plausible interpretations.

Over time, these constraints become predictive. Metadata is not merely descriptive; it is generative. It does not simply accompany content—it can often **approximate or infer it**, especially when aggregated at scale.

Search queries reveal intent. Communication frequency reveals relationship strength. Co-location reveals association. At sufficient scale, these signals converge into highly accurate behavioral models that frequently render direct content access unnecessary.

Corporate systems optimize behavior for monetization; state systems constrain it for control—but both rely on the same underlying machinery: **prediction through large-scale behavioral inference**.

IV. Identity Without Escape: Persistent Anchors

A defining feature of industrial profiling is the emergence of **persistent identity**.

Earlier systems relied on mutable identifiers—names, documents, addresses—that could be altered or obscured. Modern systems reconstruct identity through overlapping signals:

- Device fingerprints
- Behavioral patterns
- Social graphs
- Biometric markers (faces, gait, voice)

Publicly shared images serve as durable anchors. Even when individuals change accounts or adopt pseudonyms, facial recognition systems—particularly in state or intelligence contexts—can reconnect identities across datasets. Co-occurrence in photos or shared events further strengthens inferred relationships.

The implication is profound:

> Identity is no longer something one declares, but something continuously inferred.

This eliminates much of the friction that once constrained surveillance. Identification does not depend on any single signal; it emerges from redundancy across many.

V. Fusion: From Data Points to Ontologies

The culmination of this evolution is **data fusion**: the integration of disparate datasets into unified analytical systems.

Platforms such as Palantir Technologies aggregate government records, financial transactions, social media activity, location data, and communications metadata into coherent

models of individuals and networks. These systems construct dynamic ontologies that allow analysts to query relationships, detect patterns, and generate predictions.

A concrete example illustrates the shift. In immigration enforcement, Palantir's Enhanced Leads Identification and Targeting for Enforcement (ELITE) tool populates maps with potential targets, drawing on visa records, employment data, phone metadata, social connections, and even Medicaid or HHS address information to assign "address confidence scores" and generate dossiers. Officers can identify "target-rich" neighborhoods for operations, flagging individuals not solely on direct evidence but because their **behavioral and relational signature** resembles previously identified cases. Similar fusion appears in tools like ImmigrationOS, which integrates travel histories, biometrics, and social data for prioritization.

Suspicion is no longer discovered—it is **generated**.

Profiling does not merely document reality; it actively constructs it by surfacing probabilistic associations that become operationally actionable.

VI. From Explanation to Preemption

Traditional profiling was largely retrospective. It sought to explain past actions—who committed a crime, who organized a plot, who posed a threat.

Industrial profiling is predictive and preemptive. It identifies:

- Who might commit a crime
- Where crime might occur
- Who might default, radicalize, or deviate

This logic is often compared to the vision depicted in *Minority Report*, where individuals are apprehended before committing crimes. While contemporary systems lack deterministic foresight, the structural resemblance is clear: predictive policing tools analyze historical data, 911 calls, license-plate readers, and social signals to generate "heat lists" or risk scores.

Modern systems operate on probability. Individuals are flagged not because they will act, but because they **statistically resemble others who have**.

The shift is subtle but profound:

> Individuals are no longer judged primarily on actions, but on their position within a probabilistic landscape.

Suspicion becomes structural—generated continuously rather than triggered by discrete events.

VII. Law in the Age of Inference

Legal frameworks such as the General Data Protection Regulation attempt to impose limits through consent, transparency, and minimization. Yet they face structural constraints.

Most legal systems regulate **data as an object**. Modern profiling derives power from **relationships and inferences**, which are far harder to define, observe, or constrain.

Additional challenges include:

- Continuous data flows across jurisdictions
- Broad exceptions for national security and “legitimate interests”
- Opaque algorithmic systems resistant to oversight

The result is a persistent mismatch:

> Legal frameworks designed for an age of records struggle to govern an age of continuous, predictive inference.

VIII. The Asymmetry of Power

Industrial profiling produces a structural imbalance.

Individuals generate data continuously through participation in modern life. Avoidance is possible but costly and incomplete. Meanwhile:

- Corporations maintain opaque systems protected by secrecy
- States access and integrate data through legal authority or partnerships
- Technical complexity obscures accountability

The result is a clear asymmetry:

> The many are rendered legible; the powerful remain comparatively opaque.

IX. Internalization: Profiling and the Self-Regulation of Behavior

Beyond its institutional and technological dimensions, the industrialization of profiling produces a profound psychological transformation. Surveillance no longer operates solely as an external force; it becomes internalized.

This dynamic was anticipated by Michel Foucault in his analysis of the panopticon: a theoretical prison design by Jeremy Bentham in which inmates, visible to a central observer they cannot see, internalize discipline and regulate their own behavior under the uncertainty of constant watching. The power of the panopticon lies not in perpetual observation but in the **anticipation** of it.

Industrial profiling extends this logic dramatically. Individuals operate within environments where actions may be recorded, analyzed, and interpreted in opaque ways—by

platforms optimizing for engagement or states assessing risk. The result is a shift toward **self-regulation**.

This manifests as:

- Self-censorship in posts, searches, or associations
- Avoidance of certain groups, topics, or locations
- Alignment with perceived norms to minimize risk scores
- Modification of behavior across digital and physical contexts

Crucially, these adaptations do not require explicit coercion. They arise from anticipation.

> Control is exercised not only through what systems do, but through what individuals avoid doing.

The effects extend beyond individuals. As people self-censor and self-sort, the data generated reinforces patterns, shaping future predictions. The system not only observes reality—it subtly reshapes it, creating feedback loops that normalize conformity.

X. The End of Selective Surveillance

Profiling has undergone a fundamental transformation:

- From **targeted** to **universal**
- From **manual** to **automated**
- From **retrospective** to **predictive**
- From **fragmented** to **integrated**

Earlier systems were constrained by friction—cost, time, human attention. Industrial systems remove these constraints. Monitoring becomes ambient. Inclusion becomes default.

The principle that data should serve only its immediate purpose has given way to a paradigm in which **all data is potentially exploitable**.

XI. Conclusion: The Price of Participation

The long arc from postal secrecy to digital data fusion reveals a consistent pattern: each technological expansion increases the scope of profiling, while legal and social responses lag behind. What distinguishes the present is structural. Profiling is no longer an activity directed at specific individuals—it is an infrastructure within which individuals exist.

The category of "person of interest" dissolves. Everyone becomes subject to continuous evaluation.

This transformation is sustained not only by state power, but by economic incentives. Platforms that appear free operate through behavioral data extraction. The phrase *"if you're not paying for the product, you are the product"* captures an intuition—but understates the reality.

What is produced is not the individual, but a **predictive model** of the individual—portable, actionable, and often inaccessible to the person it represents.

A central challenge lies in a gap between perception and reality.

First, people underestimate the **impact** of what is known. Profiling operates through association. Relationships—past, weak, or indirect—can shape outcomes. A connection to someone who later becomes undesirable may influence opportunities. One is judged not only individually, but relationally.

Second, people underestimate the **scope** of what can be known. Systems infer sensitive attributes—political, religious, sexual, economic—not from explicit disclosure, but from patterns. These inferences become operational regardless of accuracy.

> Individuals are evaluated not only on what they reveal, but on what can be inferred—and on who they are connected to.

Participation in digital life thus entails an implicit exchange: convenience for legibility. This exchange is neither transparent nor negotiable.

The challenge is not to halt datafication, but to constrain it—to restore friction, enforce limits, and ensure accountability.

The central question is clear:

> Will intervention occur before the infrastructure of permanent profiling becomes too deeply embedded to meaningfully challenge?

Absent such intervention, the cost of participation is not merely data—but the gradual erosion of the boundary between being observed, being inferred, and ultimately being defined.